

The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins - Proteins 2009

Ceslovas Venclovas and Mindaugas Margelevicius, Institute of
Biotechnology, Vilnius, Lithuania

Anton Thomma

Technical University Munich

May 18, 2010

Overview

- 1 Human expertise in template-based modeling
- 2 HHPred and CS-Blast/CSI-Blast from Soeding Group

Steps of template-based modeling

- 1 Template detection
- 2 Template selection
- 3 Sequence-structure alignment
- 4 Model construction
- 5 Model evaluation

Each step crucial for model quality

In the process of template-based modeling each step can only be as good as the selection from the step ahead.

Basic settings and ideas

- Template-based modeling is the most accurate protein structure prediction method
- If the relationship is close → detection of related templates trivial
- Basic problem of template based modeling:
quality of model only as good as:
 - ▶ selected target
 - ▶ template-target alignment
- Provided that distantly related structures are reliably detected
 - ▶ The quality of template modeling is primarily determined by the accuracy of the alignment between the target sequence and the structural template

Input of human knowledge

Steps where human input is valuable

- Selection of the best template
- Construction of the optimal sequence-structure alignment

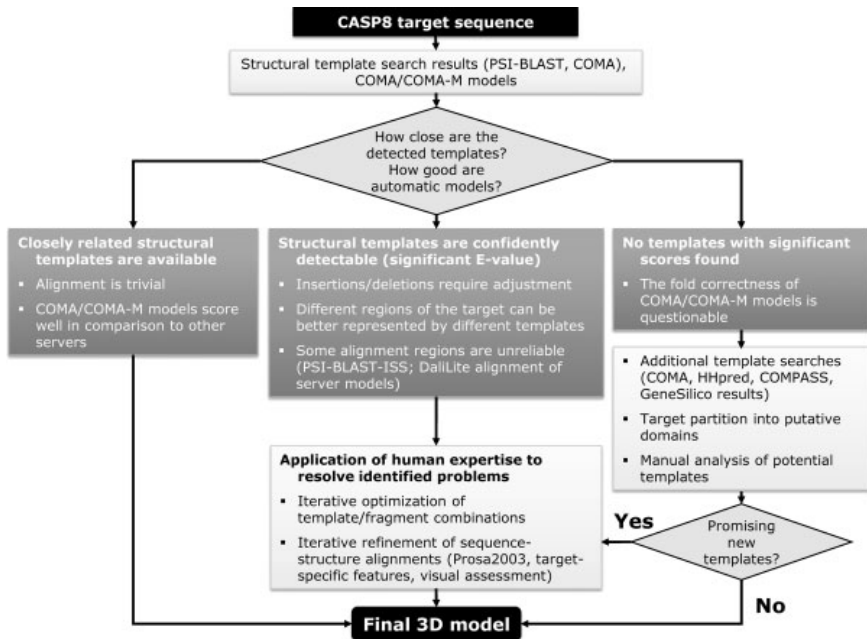


Figure: Flowchart of the human expert input - Venclovas C. - Proteins 2009

Sequence-structure alignment

- Reliably aligned regions were first identified with PSI-BLAST-ISS
- PSI-BLAST-ISS performs transitive searches and validates the template selection
- Automatic server models from other CASP 8 groups were superimposed with representative templates using DaliLite
- Many alignment errors affect β -strands that are at the edge of β -sheets
- The N-terminal region of T0413 displayed no consensus alignment
- The Pro residue can be accommodated without interfering with main chain hydrogen bonding

Sequence-structure alignment with expert input

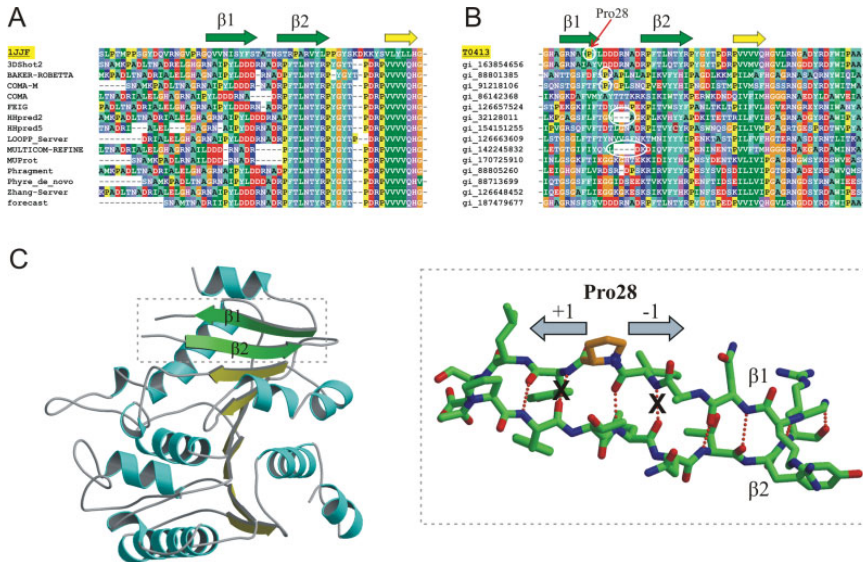


Figure: Resolving sequence-structure alignment uncertainty for the edge β -strand - Venclovas C. - Proteins 2009

Overview

1 Human expertise in template-based modeling

2 HHPred and CS-Blast/CSI-Blast from Soeding Group

Protocol of HHpred servers

HHpred

Designed as an interactive function and structure prediction server

- HHpred is based on template search with up to eight PSI-Blast searches
- A profile hidden Markov model is calculated from the target alignment and used for the search of homologous templates
- Rank target-template alignments and select the template most closely related to the target
- Run MODELLER with standard automodel script with the template-target multiple alignment
- Little run time with no alternative alignments, no side chain optimization, no contacts predicted, no loops modeled and no structural clustering

HHPred in CASP 8

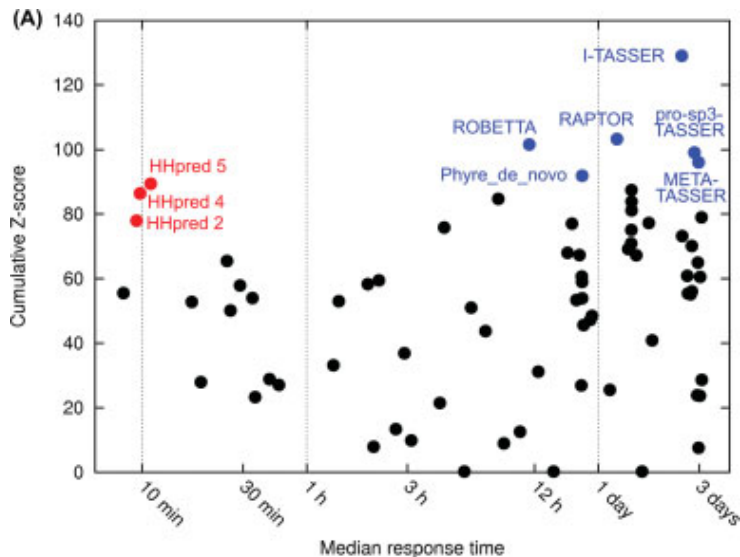


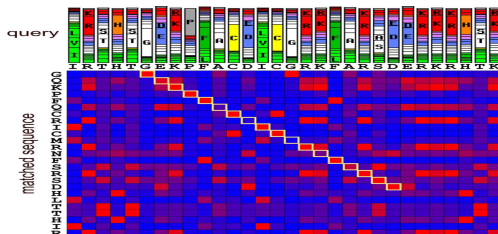
Figure: Model accuracy against response time - Hildebrand, A and Söding, J.
- Proteins 2009

CS/CSI-Blast - Context-specific BLAST

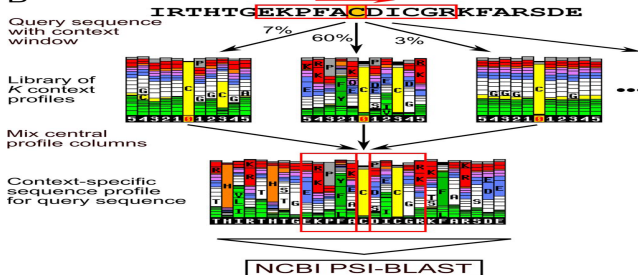
- BLAST calculates the total similarity score with "substitution matrices"
- So the similarity between amino acids is modeled independent of the "sequence context"
- Context-specific amino acid similarities: a residue's context can say a lot about what amino acids it is likely to mutate into
- Context-specific BLAST is an extension of BLAST that uses context-specific substitution scores
- CS-BLAST jump-starts BLAST with the sequence profile from the query sequence

CS/CSI-Blast - Context-specific BLAST

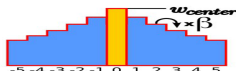
A



B



C



Performance of CS-BLAST

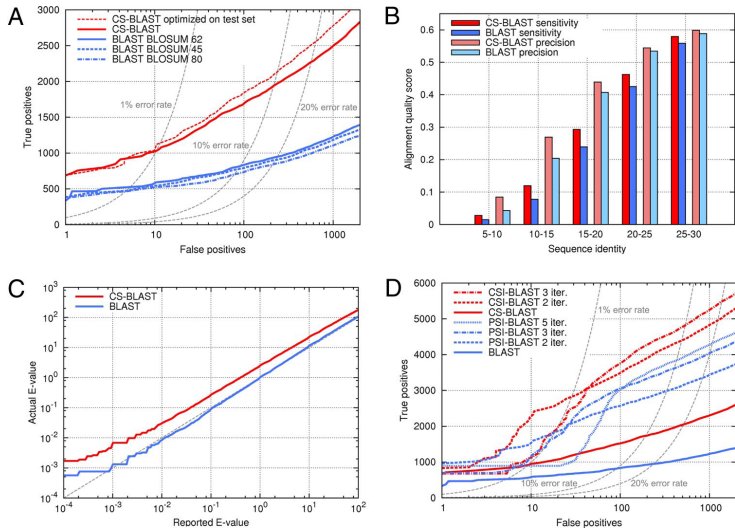






Figure: Context information improves search performance and alignment quality - Söding J. - Biophysics and Computational Biology 2009

References

-  A. Biegert and J. Söding, *Sequence context-specific profiles for homology searching*, Proceedings of the National Academy of Sciences **106** (2009), no. 10, 3770–3775.
-  A. Hildebrand and J. Söding, *Fast and accurate automatic structure prediction with HHpred*, Proteins: Structure, Function, and Bioinformatics **77** (2009), no. S9, 128–132.
-  M. Margelevicius and C. Venclovas, *Psi-blast-iss: an intermediate sequence search tool for estimation of the position-specific alignment reliability*, BMC Bioinformatics **6** (2005), no. 1, 185.
-  _____, *The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins*, Proteins: Structure, Function, and Bioinformatics **77** (2009), no. S9, 81–88.